# Foundational Ontologies – What for?
# Motivations for SUMO and DOLCE

**Bachelorarbeit**

zur Erlangung des akademischen Grades

Bachelor of Arts (B.A.)

im Fach Bibliotheks- und Informationswissenschaft

Humboldt-Universität zu Berlin

Philosophische Fakultät I

Institut für Bibliotheks- und Informationswissenschaft

eingereicht von Anja Eisemann

geb. am 03.09.1982

in Hamm / Westfalen

1. Gutachter: Prof. Dr. Stefan Gradmann

2. Gutachterin: Prof. Vivien Petras, Ph.D.

Berlin, den 04.11.2009

**Content**

**List of Figures**



**List of Tables**

# 1. Introduction

## 1.1 Research question and research objective

As library and information scientists we are familiar with organising information, words and concepts. Classification is one of our core activities. We use thesauri and controlled vocabularies. But what about ontologies? Ontologies are means of organising concepts as well. Ontologies are one of the basic components of the Semantic Web[1] as described by Berners-Lee et al. (2001). Should we not be concerned about ontologies, too?

Mainly there are two types of ontologies: foundational and domain ontologies. While there may be an unlimited number of domain ontologies – one for every thinkable domain –, there are only few foundational ontologies, because they aim at describing the basic structure of everything. (In theory only one is feasible.) Being in my optinion the more peculiar ones, this thesis is about foundational ontologies. My research question is: "What use do foundational ontologies have?"

By answering this question I hope to provide library and information scientists with a basic understanding of what foundational ontologies are, and where it is meaningful to apply them.

## 1.2 Outline

This paper is structured into five parts. In the introduction, I present my research question and motivation and an outline of this paper. Further I will explain what foundational ontologies are by giving an overview of common definitions. Afterwards I will introduce the two foundational ontologies that I am going to investigate: the Suggested Upper Merged Ontology (SUMO) and the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE).

The following three chapters form the research part of this paper. Chapter 2 gives a more thorough insight into SUMO and DOLCE by comparing their basic ontological choices. Chapter 3 focuses on the motivations for SUMO and DOLCE as expressed by their authors. Chapter 4 presents a first overview about what SUMO and DOLCE have been used to so far, based on a citation analysis of the main publications about SUMO and DOLCE.

The findings of the paper are summarized in Chapter 5.

1 The Semantic Web provides concepts of the future Internet that will greatly improve information exchange in that it makes its content understandable and processable to machines and artificial agents.

### 1.3 Definitions

Preceding the following definitions I would like to point out that it is my approach to present authoritative definitions and chose one that is most suitable for the purpose of this paper. It is *not* my approach to look into the actual use of the defined words or to elaborate a new definition. It is important to keep in mind that is commonly understood to be an ontology is not necessarily conform to what will be formally defined as an ontology in this paper.

#### 1.3.1 Ontology (Philosophy)

The term "ontology" appears in philosophical writings of the 17th century for the first time. It is derived from Greek and can roughly be translated as "the study of being / of that which is" (λόγος *logos* – science, study, theory; ὄντος *ontos* neuter participle genitive of εἶναι *einai* – to be). Although not always called ontology, the study of that which there is has a long tradition in European philosophy. It dates back to what was called "first philosophy" by Aristotle and "metaphysics" by his successors. (Hügli & Lübcke 2001, p. 468). Meaning a certain philosophical discipline, the term "ontology" is used as an uncountable noun, and usually written with a capital 'O'.

#### 1.3.2 Ontology (Artificial Intelligence)

Since the 1980s the term "ontology" has also been used in Artificial Intelligence (AI) research. (Gruber 2007). The most cited, and also most debated, definition in this context originates from Gruber (1993). He states that an ontology is "*an explicit specification of a conceptualization*"[2] (p. 199). By conceptualization he means "an abstract, simplified view of the world, that we wish to represent for some purpose" (p. 199). The notion of "conceptualization" in Gruber's definition – as well as the use of the word "ontology" by different authors – has been discussed in detail by Guarino & Giaretta (1995), resulting in a more detailed definition of "ontology" and related terms. According to Guarino & Giaretta (1995) the word ontology is used in two senses, either meaning "*a logical theory which gives an explicit, partial account of a conceptualization*" or as a *synonym of conceptualization*. A conceptualization in turn is "an intensional semantic structure which encodes the implicit rules constraining the structure of a piece of reality"

2 This definition as well as the following ones are set in italics by me.

(Guarino & Giaretta 1995). This definition, as well as Gruber's original one, seems to be exclusively targeted at an audience with substantial background in AI.

Although much debated, and biased towards AI, Gruber's (1993) definition is cited by many recent publications that introduce ontologies. Staab & Studer (2004) cite the following adapted definition in the preface to their Handbook on Ontologies: "*An ontology is a formal explicit specification of a shared conceptualization for a domain of interest*" (p. VII). Thus they claim that ontologies can only be developed for a "limited domain and a limited group of people" (p. VIII), and stress that an ontology has to be "specified in a language that comes with formal semantics" (p. VII).

Schneider (2008) gives a similarly adapted definition in the Handbook of Knowledge Representation: "*An ontology is an explicit specification of a shared conceptualization that holds in a particular context*" (p. 936). While not making any claims on the specification language, this definition, too, emphasises that the area of application for one ontology is limited to one domain – one "viewpoint" or "context" (p. 937). This confinement presents a problem when talking about foundational ontologies (see chapter 1.3.4).

The obvious problem, however, with the definitions listed so far is that they are not very helpful to people who are new to the field. Stuckenschmidt (2009) therefore decides explicitly in his German-language introduction to ontologies that he won't give a definition. Additionally, the existing definitions are either to vague for his taste, like Gruber's, or too strict, like the refined definition by Staab & Studer. (p. 22-23).

### 1.3.3 Ontology (Information Science)

A different approach arises when viewing ontologies from the perspective of library and information science. As done on the website of the Standard Upper Ontology effort of the IEEE[3], it is understood that ontologies are "similar to a dictionary or glossary" (IEEE 2003, December 28), but have additional features available.

This approach has been taken by Stock & Stock (2008). They, too, cite an amended version of Gruber's (1993) unavoidable definition: "*An ontology is a formal, explicit specification of a shared conceptualization*" (p. 255) but dismiss it as too broad. Elaborating on a distinction between "lightweight" and "heavyweight" ontologies

3 The IEEE (read: I-triple-E) is a non-profit organisation for the advancement of technology. It is an important scientific publisher and, through the IEEE Standards Association, an important developer of international industry standards. (http://www.ieee.org/).

introduced by Corcho et al. (2003, p. 44), they assert that lightweight ontologies include controlled vocabularies, classifications, thesauri, and more; whereas heavyweight ontologies have to fulfill the following four additional criteria (Stock & Stock p. 256-257):

(1.) use of a standardised ontology language

(2.) allow for automated reasoning by use of terminological logic[4]

(3.) occurrence of and distinction between general and individual terms

(4.) use of specific relations (apart from hierarchical relations)

Therefore Stock & Stock's (2008) definition in reads as follows:

"Unter einer Ontologie (i.e.S.) verstehen wir eine Begriffsordnung, die in einer standardisierten Sprache vorliegt, automatisches Schlussfolgern gestattet, stets über Allgemein- und Individualbegriffe verfügt sowie neben den Hierarchierelationen mit weiteren spezifischen Relationen arbeitet."[5] (p. 269).

I prefer the latter reading of their definitition that is open about which logic can be used for reasoning.

The definition by Stock & Stock is to a large extend compatible with a newer definition by Gruber (2007):

"In the context of computer and information sciences, an ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application."

The main difference between these two definitions is that Gruber doesn't seem to count individual terms as part of the ontology proper. This view is not uncommon; for example Guarino (1998) discerns the ontology, containing state-independent information, from the core knowledge base, containing state-dependent information. Stock & Stock (2008) make the same distinction, but within the ontology (compare p. 264-265).

I find Stock & Stock's definition most useful for an introduction to the topic because it is easiest to understand for LIS professionals and it allows you to check if something is an ontology or not easily. The question whether assertions about state-dependent

4 A terminological logic is based on semantic networks and operates on concepts and relationships. (compare Stock & Stock 2008, p. 260-264.)

5 My translation: "An ontology is an order of terms *(Begriffsordnung)* that comes in a standardised language, allows for automated reasoning, commands general and individual terms, and uses other specific relations apart from hierarchical relations."

information should be considered a separate part of the ontology or not at all a part of the ontology proper appears to be less relevant for a definition of the core concept.

### 1.3.4 Foundational ontologies

As said before, ontologies can roughly be divided into two types. The majority of ontologies are so-called *domain* ontologies, meaning that they are intended to model the concepts and relations that hold in a particular domain of interest. A lot of such ontologies exist for example in the field of biomedicine. (Schreiber 2008, p. 939).

The other type of ontologies are called *foundational* ontologies (also: *upper* ontologies, *top-level* ontologies or similar). These ontologies aim to describe "general notions, such as time, space, events and processes" (Schreiber 2008, p. 939). Gruber (1993) employs a nice image to depict foundational – or, as he calls them, "comprehensive content ontologies": They function like "a 'conceptual coat rack' on which to 'hang' more specific ontologies and domain knowledge" (p. 215). Examples of foundational ontologies are SUMO and DOLCE.

A more granulated classification of ontologies has been proposed by Oberle (2006, p. 43-47). He classifies ontologies according to their purpose, specificity and expressiveness.

Different ***purposes*** are served by application and reference ontologies. *Reference* ontologies are used during development time of applications for establishing consensus on the meaning of terms. *Application* ontologies are used during run time of a specific application, they have therefore to use a decidable logic. Application ontologies may describe specific worlds.

Lightweight and heavyweight ontologies differ with respect to their ***expressiveness***. *Heavyweight* ontologies specify the intended meaning of a vocabulary as precisely as possible, using a lot of axioms and an expressive language. *Lightweight* ontologies are much less expressive and intended to be shared among users that have already formed a common understanding of the vocabulary.

Three layers of ***specificity*** can be discerned in ontologies. *Generic* ontologies define concepts that are generic across many fields (such as state, process, component, etc.). *Domain* ontologies define concepts that are specific for a certain domain. These concepts are often specializations of concepts in the generic ontologies. *Core* ontologies are placed in the middle between generic and domain ontologies. They define concepts that are generic across a set of domains. There are no clearly defined borders between the three layers of specificity.

Finally, Oberle (2006) characterizes foundational ontologies as generic, reference and heavyweight ontologies. (p. 48, compare Fig. 1).

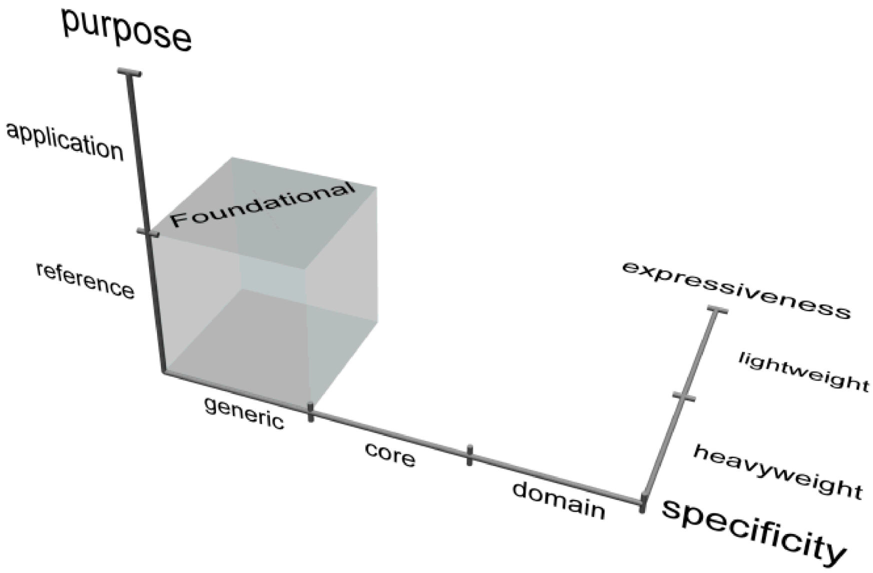


*Figure 1: Classification of ontologies and the role of foundational ontologies (Oberle 2006, p. 49)*

### 1.3.5 Ontology languages

The definition of an ontology requires that a standardised ontology language is used.

There does not seem to be an agreed-upon definition of "ontology language",[6] although a variety of languages are labeled as such. The spectrum includes conceptual graphs (mentioned by Schneider 2008, p. 936), F-logic (Angele & Lausen 2004), description logics based languages (Baader et al. 2004) and quantified modal logic (Oberle 2006, p. 41-42).

According to Passin (2004), ontology languages are the frameworks for constructing ontologies. They typically provide a syntax, a vocabulary, and some predefined terms. And they may be used to specify rules to support certain logical inferences. (p. 151). So, in my perception, ontology languages provide the basic semantics that enable a logical reasoner to reason on an ontology's content. (The non-basic semantics can be constructed using the chosen ontology language.)

6 Especially the distinction between knowledge representation languages and ontology languages proves difficult. (compare Schneider 2008, p. 237-238).

The best-known ontology language is the *Web Ontology Language (OWL)*. It was created by the W3C Web-Ontology working group in order to serve as the standard ontology language of the Semantic Web. (Antoniou & Harmelen 2004, p. 67). OWL has got the status of a W3C recommendation. A guide can be found on the W3C's website[7].

OWL uses the RDF/XML syntax. It comes in three 'flavours' to account for different requirements with regard to expressiveness and computability. OWL Full is based on RDF Schema[8] and quite expressive, but undecidable by formal reasoning. OWL DL (for description logic) is a sublanguage of OWL Full and corresponding to a well studied description logic. It thus allows for efficient reasoning support. OWL Lite in turn is a sublanguage of OWL DL with even more restricted expressivity but the advantage of easy use and implementation. (Antoniou & Harmelen 2004, p. 70-71).

OWL allows to define classes, properties and individuals. The possibility to use URIs as object values makes OWL suitable for use in the web. (Stock & Stock 2008, p. 267).

Another language that is often used with foundational ontologies is called the *Knowledge Interchange Format (KIF)*. KIF was developed at the Computer Science Department of Stanford University as part of the DARPA Knowledge Sharing Effort[9]. The specification of KIF was written by Genesereth & Fikes (1992) and is also available from the KIF website[10].

KIF "provides for the expression of arbitrary sentences in predicate calculus [i.e. first order logic]" (Genesereth & Fikes 1992, p. 5), so it is much more expressive than even OWL Full.

Foundational ontologies need to be written in a language like KIF. Using OWL alone would be "non-sensical" (Masolo et al. 2003, p. 6) because it doesn't allow them to characterise their concepts in suitable detail.

7 http://www.w3.org/TR/owl-guide/
8 RDF stands for Resource Description Framework. RDF and RDF Schema are explained in McBride (2004).
9 An overview of the DARPA Knowledge Sharing Effort is available online at http://www-ksl.stanford.edu/knowledge-sharing/papers/kse-overview.html .
10 http://www.ksl.stanford.edu/knowledge-sharing/kif/

### 1.4 Introducing SUMO and DOLCE

#### 1.4.1 Why SUMO and DOLCE?

Next to Cyc (http://www.cyc.com/), SUMO and DOLCE are the most prominent and advanced examples of foundational ontologies.[11] While Cyc is by far the oldest ontology project, only part of it (called OpenCyc, http://opencyc.org/) is released under a free license. Both SUMO and DOLCE are freely available, and therefore chosen as the two example ontologies for this paper.

#### 1.4.2 SUMO

SUMO stands for Suggested Upper Merged Ontology. SUMO is owned by the IEEE but made freely available[12]. The homepage for SUMO is http://www.ontologyportal.org/ , and its technical editor is Adam Pease. As of October 2009, and including its associate domain ontologies, it contains 20,000 terms and 70,000 axioms. Considering only SUMO proper, the terms number about 1000 and the axioms 4000. (Mascardi et al. 2007).

SUMO was proposed as a starter document for the Standard Upper Ontology (SUO) effort in 2001. The IEEE-sanctioned SUO working group aimed to develop one large, public ontology that should provide "definitions for general-purpose terms, and [...] act as a foundation for more specific domain ontologies" (Niles & Pease 2001).[13]

The knowlegde representation language used by the project is a simplified version of KIF, called SUO-KIF. (Niles & Pease 2001). There also is an OWL Full version[14]. (Oberle 2006, p. 105).

SUMO was built by merging the content of several upper ontologies that were freely available on the web. After gathering the ontological content, all of it was translated to

11 When introducing foundational ontologies, Schreiber (2008, p. 939) mentions SUMO and DOLCE, Euzenat & Shvaiko (2007, p. 68) mention Cyc, SUMO, and DOLCE.

12 This means that SUMO comes with a license that states the IEEE as the sole copyright holder, but grants everyone "a perpetual, non-exclusive, royalty-free, world-wide right and license to copy, publish and distribute the Document in any way, and to prepare derivative works that are based on or incorporate all or part of the Document provided that the IEEE is appropriately acknowledged as the source and copyright owner in each and every use." (Pease 2008, November).

13 This working group has ceased to exist by now, compare this email by Jim Schoening to the SUO mailing list: http://grouper.ieee.org/groups/suo/email/msg13629.html .

14 The OWL-version of SUMO contains the following warning: "this is a very lossy translation of the original KIF version. OWL does not have the expressive capability of KIF and much of the content of SUMO is lost in translation. Please look at the KIF source or online browser at http://www.ontologyportal.org for the full version." (http://www.ontologyportal.org/translations/SUMO.owl.txt).

SUO-KIF. Then two ontologies were identified that defined very high-level concepts. These two were merged, forming a foundation for aligning the rest of the content. During the alignment process, the authors of the ontology had to fill in gaps between concepts, sort out notions that did not fit with the rest of the ontology and find solutions to cases of partially overlapping concepts. (Niles & Pease 2001).

The basic structure of SUMO is depicted in Figure 2.

The Structural Ontology consists of definitions of certain syntactic abbreviations on the basis of the vocabulary that SUO-KIF provides. The abbreviations facilitate computation and understanding of the ontology. The Base Ontoloy contains the hierarchy of the highest-level concepts. Together with the Set/Class Theory section, the Numeric section, the Temporal section, the Mereotopology section, the Graph section, the Measure section, the Processes section, the Objects section, and the Qualities section, these are forming SUMO proper.

The Mid-Level Ontology (MILO) is an addition to SUMO which connects several domain ontologies to SUMO. (Stuckenschmidt 2009, 81).

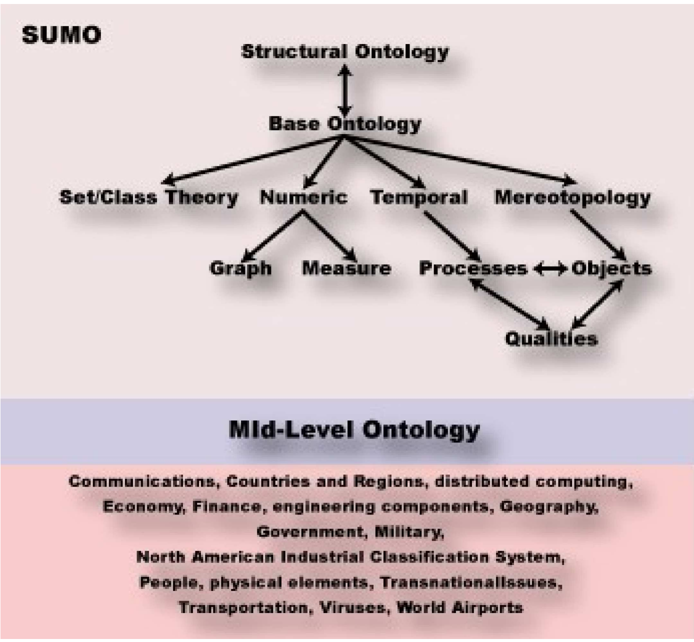


*Figure 2: The basic structure of SUMO (Pease 2009, August 11)*

- Entity
  - Physical
    - Object
    - ContentBearingPhysical
    - Process
    - PhysicalSystem
  - Abstract
    - Quantity
    - Attribute
    - SetOrClass
    - Relation
    - Proposition
    - Graph
    - GraphElement

*Figure 3: Concept hierarchy of SUMO v1.81 (compare Pease 2008, November)*

Since its beginning, there have been quite a lot of changes in SUMO's concept hierarchy, even up to the third level. The current version of SUMO is 1.81, and Figure 3 displays the first three levels of its concept hierarchy.

#### 1.4.3 DOLCE

DOLCE stands for Descriptive Ontology for Linguistic and Cognitive Engineering. DOLCE was developed at the Laboratory for Applied Ontology in Trento, Italy, as part of the WonderWeb Foundational Ontologies Library (WFOL). The WonderWeb project received fundings under the Fifth Framework Programme of the European Union from 2002 to 2004. All versions and documentation of DOLCE are freely available at http://www.loa-cnr.it/DOLCE.html. The last update was in June 2006. In contrast to SUMO, DOLCE was not build as a one-fits-all foundational ontology, but rather intended to serve "as a reference module, to be adopted as a starting point for comparing and elucidating the relationships with other future modules of the [WFOL] library" (Masolo et al. 2003, p. 5).

DOLCE is written in the KIF language and also translated – as far as possible – to DAML+OIL and OWL DL. (Oberle 2006, p. 101). It contains about 100 concepts, and a similar number of axioms. (Mascardi 2007).

As the name suggests, DOLCE has been built reflecting human language and cognition. Considering its intended purpose, naturally all basic ontological choices for DOLCE were made deliberately and are well documented.

DOLCE is a single ontology, and is not divided into modules.[15] The basic categories of DOLCE are depicted in Figure 4.

This figure is juxtaposed with Figure 1, the basic categories of SUMO. The most significant differences will be specified in chapter 2.

15 However, additional theories are available and can be included on demand. Oberle (2006) mentions Descriptions and Situations, the Ontology of Plans, the Ontology of Time, and the Ontology of Information Objects. (p. 100).

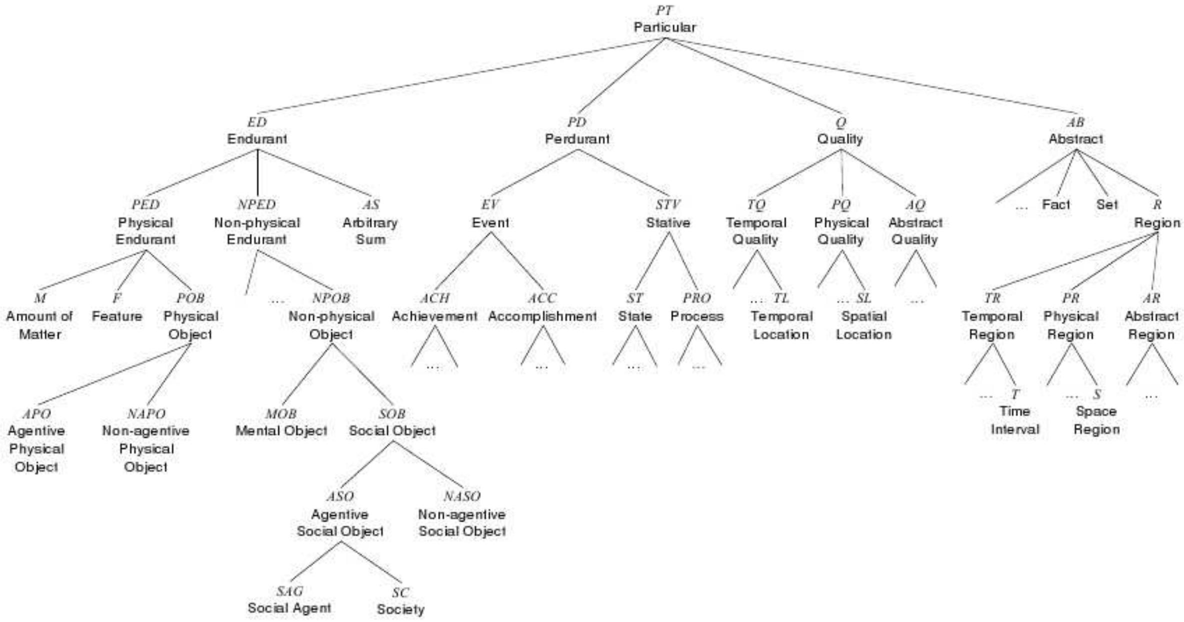


*Figure 4: DOLCE basic categories (Masolo et al. 2003, p. 14).*

## 2. What are SUMO's and DOLCE's basic ontological choices?

In order to get better acquainted with SUMO and DOLCE, and to know their fundamental differences, it is helpful to compare their respective modelling approaches and commitment to different world-views, a.k.a. basic ontological choices.

### 2.1 Methodology

The final project report of the WonderWeb Foundational Ontologies Library (Masolo et al. 2003) describes a number of design problems and ontological choices that have to be considered when creating a foundational ontology. First I present these options. Then I will analyse the literature (and documentation) available about SUMO and DOLCE in order to identify their respective choices.

### 2.2 Basic ontological choices

When creating a foundational ontology, either a *descriptive* or a *revisionist* approach can be taken. If you use natural language and cognition as the basis for your ontology, this approach is called descriptive. Human commonsense, for example, regards things (spatial objects) and events (things that happen in time) as two distinct entities. A scientist or philosopher, on the other hand, might argue that time is nothing but the fourth dimension of an object and thus there is no 'intrinsic' difference between things and events. The approach to model the world independently from the way that humans conceptualise it is called revisionist. (Masolo et al. 2003, p. 7).

An ontology that employs only a minimal number of primitives (i.e. basic concepts) to model all the remaining concepts is called *reductionist*. An ontology that aims at maximal expressivity, using a large number of basic concepts, is called *multiplicative*. The problem of co-localized entities can point out the difference between these two approaches: Take the example of a vase that is made from an amount of clay. Are the clay and the vase two different entities, although co-localized in space and in time? A multiplicativist's answer would be yes. In a reductionist ontology on the other hand there can only be one entity at each space-time location. The reductionist ontology would describe the differences between the clay and the vase as different views of the same spatio-temporal object. (Masolo et al. 2003, p. 7-8).

The basic assumption of *actualism* is that only what is real exists. *Possibilism* on the other hand admits also possible entities. The adoptance of possibilism goes hand in hand with either employing a language that comes with modal operators, or modelling modality by quantifying over situations or worlds.[16] In the first case one can translate the expression "It is possible that John is ill" in a literal fashion. In the other approach, one has to rephrase the expression into "There is a world in which John is ill". This latter sentence can be translated literally.
Generally, the problem occurrs the other way round: Deploying a first-order-logic often determines a preference towards actualism. (Masolo et al. 2003, p. 8).

In the field of ontologies we distinguish between *universals* (entities that can have instances) and *particulars* (entities that can not have instances). The distinction between universals and particulars is similar to that between common nouns and proper nouns. The common noun 'cat', refering to an universal, can have instances, like the little black cat Audrey, refering to a particular. Audrey herself is an instance of 'cat' and cannot have

16 Similarly, presentism claims that only what is present exists, while eternalism assumes that the past and the future also exist. (Masolo et al. 2003, p. 8). I will consider this choice no further because it is highly unlikely for a foundational ontology not to model such basic notions such as change or process.

instances herself.
Properties and relations are usually considered as universals, while some entities (sets, predicates, abstracts) are problematic in their characterization as such. The designer of an ontology has to choose whether to include universals in the domain of discourse or not[17], and how to treat sets, predicates and abstracts. (Masolo et al. 2003, p. 9)

The objects in an ontology can be characterized as either *abstract* or *concrete* entities. The most common definition of abstracts is that these are entities that exist neither in space nor in time. Concrete entities are at least located in time.[18] (Masolo et al. 2003, 10). This distinction in ontology design does not require a simple either-or-choice, but rather a decision on which entities to model as abstracts and which as concretes.

Another fundamental ontological distinction can be drawn between a *3D* and a *4D* view of the world. When adopting the 3D view, objects are claimed to extend in three-dimensional space, and to be wholly present at each instant of their being. From a 4D point of view, objects are extended into space and time, and only partially present at each instant. (p. 10). It is important to note that while a 3D ontology naturally distinguishes between objects and processes (Niles & Pease 2001, p. 5), this distinction is usually not present in a 4D ontology, but may be modelled in 4D also. (Masolo et al. 2003, p. 11).
A descriptive and multiplicative approach is likely to go hand in hand with the 3D view, while a revisionist and reductionist approach tends towards a 4D view.
The issue of 3D vs. 4D is closely related to the notion of *endurants* and *perdurants*. Endurants are entities that are wholly present at any time of their existance, while perdurants consist of different temporal parts. Put simply, endurants correspond to 3D objects and perdurants correspond to 4D objects.[19]

17 Note that even when universals are excluded from the domain of quantification, they do occur indirectly to organise and structure the domain. (Borgo et al. 2002, p. 12).
18 This definition raises the question of how something can exist without existing in space or time. An alternative definition is that abstracts possess no casual power while conctreres do. That definition is problematic as well because we can think of, for example, 'the center of mass of the solar system' which has got no casual power but we still wouldn't like to count as an abstract. Masolo et al. (2003) chose to adhere to the first definition. (p. 10).
19 Although listed as two different subsections in (Masolo et al. 2003), there doesn't seem to be any difference between the 3D/4D opposition and endurantism/perdurantism. Semy et al. (2004) state that this conflation is "problematic" (3-9), but for their purpose, namely identifying a suitable upper ontology to be used within military and government, they decide that they can equate 3D with

I would like to emphasize that my understanding of the 3D vs. 4D paradigm is based on Masolo et al. (2003) as well as Niles & Pease (2001). They focus on how *objects* are modelled (namely either in 3D or in 4D). The modelling of *processes* is, though related, a different topic. In my understanding, an ontology that models endurants and perdurants would be 3D. An ontology that only models perdurants would be 4D. An ontology that only models endurants would be 3D as well, but probably not a foundational one. Others show a slightly different understanding or use of the terms, that can be confusing. They characterise ontologies as "4D", if they model perdurants at all. (compare Oberle et al. 2007, p. 159; Semy et al. 2004, p. 28).

### 2.3 Basic ontological choices in SUMO

SUMO is neither explicitly descriptive nor revisionist. Oberle et al. (2007) take the "common sense distinction between objects and processes" (p. 159) as sufficient to classify it as descriptive.

Borgo et al. (2002) found "that SUMO does not clearly adopt neither a multiplicative nor a reductionist approach". (sic, p. 10). Oberle et al. (2007) agree, but point out that "the major part of its theories commits to a multiplicative stance." (p. 159). Semy et al. (2004), however, claim that "[g]iven its development philosophy, SUMO is clearly multiplicative in nature". (p. 4-2).

Pease and Niles (2002) state that "there are currently no modal notions" in SUMO. Oberle et al. (2007) state that it is "unclear" (p. 160) whether SUMO adopts actualism or possibilism. The lack of modal logic in any case suggests that SUMO tends towards actualism. (compare Masolo et al. 2003, p. 8; Oberle et al. 2007, p. 158).

The statement by Borgo et al. (2002), that SUMO does not classify universals (p. 12), must be outdated because Pease (2006) explains that SUMO classifies a number of properties (i.e. universals) as well. This facilitates engineering, "as it allows common features like transitivity to be applied to a set of properties, with an axiom that is written

endurantism and 4D with perdurantism. Niles and Pease (2001, p. 5) as well as Oberle et al. (2007, p.160) equate 3D-objects with endurants and 4D-objects with perdurants without further consideration. For the purpose of this paper I will equate the 3D/4D opposition with the opposition between endurantism and perdurantism.

once and inherited by those properties, rather than having to be rewritten, specific to each property." (p. 2). Therefore SUMO is an ontology of both particulars and universals.

SUMO adopted a 3D orientation because this allowed for an easy incorporation of content from several ontologies that were 3D oriented. (Niles & Pease 2001).

In SUMO, the distinction between physical and abstract is similar to that between concretes and abstracts. Physicals seem to be a specialization of abstracts, as they necessarily exist in time and space. (Borgo et al. 2002, p. 13). Abstracts are entities that exist neither in time nor in space. SUMO models several concepts as abstracts (see Figure 2).

**2.4 Basic ontological choices in DOLCE**

DOLCE is a descriptive ontology, as dictated by its cognitive bias. The categories in DOLCE are "*descriptive* notions that assist in making *already formed* conceptualizations explicit" (Masolo et al. 2003, p. 13).

DOLCE adopts the multiplicative approach. Again, it is language use that directed this choice: We say, for instance, that a vase has a handle, but not that a piece of clay has a handle. This indicates that we take a vase and the clay it consists of as two genuinely different entities. So, DOLCE allows for different entities to be co-located in the same space-time. There are other cases, too, where two categories could possibly be reduced to each other, and DOLCE prefers to "assume the existence of both kind of entities, in order to study their formal relationships [...] rather than committing to just one kind of entity in advance" (p. 14).

DOLCE uses "the simplest quantified modal logic, namely S5 plus the Barcan Formula" (p. 26). This means that it assumes a possibilist view.

DOLCE is an ontology of particulars. The DOLCE developers take the ontology of universals as formally separated from that of particulars. An ontology of unary universals has been presented by Guarino & Welty (2000). (Masolo et al. 2003, p. 15).

DOLCE models both endurants and perdurants, and their distinction is fundamental. The main relation between them is that of participation. For example, a person, which is an endurant, may participate in a discussion, which is a perdurant. (Masolo et al. 2003, p. 15-16). This means that DOLCE's basic orientation is 3D.

Although more examples of abstract entities (sets and facts) are indicated in the figure of its basic categories, the present versions of DOLCE only considers quality regions. (p. 19). DOLCE draws a distinction between qualities and qualia, which is inspired by trope theory. According to the authors, "natural language – in certain constructs – often seems to make a similar distinction". (p. 17). Each quality in DOLCE is of a certain type (e.g. color) and has a certain value (e.g. a certain shade of red). This value is also called "quale" and describes the position of a certain quality within a conceptual space. Colors are ususally associated to a topological 2D space, and lengths to a metric linear space. So, while qualities exist (and stay the same) as long as the entities exist that they inhere to, their qualia can change. DOLCE considers space and time locations as special qualities. (p. 16-18). Perdurants have a direct temporal location. Endurants inherit the temporal location of the perdurants they participate in. Endurants and perdurants as well as temporal and physical qualities are concrete entities in DOLCE. (Borgo et al. 2002, p. 12-13).

### 2.5 Comparison

Comparing SUMO and DOLCE with regard to their basic ontological choices proves a difficult task, because most of SUMO's positions are simply not well-defined, hence the interpretation of their stance varies from author to author.

An overview of SUMO's and DOLCE's choices is composed in Table 1. Comparing their choices to those of other foundational ontologies (Table 2[20]), it shows that despite their different appoaches SUMO and DOLCE are quite similar.

20 Note that SUMO and DOLCE are labeled as 4D because of a different understanding of terms. (compare chapter 2.2).

| | SUMO | DOLCE |
|---|---|---|
| descriptive / revisionist | unclear (tends towards descriptism) | descriptive |
| reductionist / multiplicative | unclear (tends towards multiplicativism) | multiplicative |
| actualism / possibilism | unclear (tends towards actualism) | possibilism |
| universals included? | yes | only particulars |
| endurantism (3D) / perdurantism (4D) | 3D | 3D |

*Table 1: Comparison of SUMO and DOLCE ontological choices*

| Requirement \ Alternative | BFO | DOLCE | OCHRE | OpenCyc | SUMO |
|---|---|---|---|---|---|
| Descriptive | - | × | - | × | × |
| Multiplicative | - | × | unclear | unclear | × |
| Actualism | × | - | - | unclear | unclear |
| 4D | × | × | - | unclear | × |

*Table 2: Basic ontologic choices for several foundational ontologies (Oberle et al. 2007, p. 159).*

## 3. What are the motivations for SUMO and DOLCE?

In the introduction and in the previous chapter we have seen that the approaches of constructing SUMO and DOLCE were quite different, and that it can be explained chiefly by their different goals. We will now have a closer look at what motivated the development of SUMO and DOLCE.

### 3.1 Methodology

First I will present an analysis of the papers that introduced SUMO and DOLCE to a wider audience. For SUMO these papers are those that are labelled as "the main SUMO publications" on Ontologyportal (Pease 2008, August 22). That is, the SUMO paper "Towards a Standard Upper Ontology" (Niles & Pease 2001), the SUMO and Semantic Web paper "The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications" (Pease et al. 2002), and the progress report "IEEE Standard Upper Ontology: A Progress Report" (Pease & Niles 2002).

For DOLCE these papers are "Sweetening Ontologies with DOLCE" (Gangemi et al. 2002) because it presents DOLCE "in an intuitive way" (Gangemi et al. 2002), and the final report about the WonderWeb Foundational Ontologies Library "WonderWeb Deliverable

D18 : Ontology Library (final)" (Masolo et al. 2003), which contains a detailed description of DOLCE.

I will look for statements that refer to or allow conclusions about the intended purposes of SUMO and DOLCE respectively. Later I will contrast my findings about SUMO and DOLCE.

**3.2 Motivations for SUMO**

The SUMO paper:

In the introduction the need for a "comprehensive, formal ontology" , acting as a "free, public standard" is described. Application areas that would profit from SUO are web searching, automated language understanding, ecommerce, and software applications in general. The Standard Upper Ontology (SUO) being developed should meet those needs. The following purposes of SUO are named in the introduction:

- Knowledge bases and databases that use a common vocabulary provided by SUO gain some degree of interoperability.
- Already existing databases can be integrated by being mapped just once to a common ontology.
- Domain-specific ontologies that reuse terms provided by SUO can interoperate to some degree.

Chapter 2 describes how SUMO was created by merging "all high-level ontological content that did not have licensing restrictions". It is mentioned that not all available concepts were included in SUMO. An example of a philosophical concept is given, whose inclusion was not deemed justified "in an engineering-oriented context".

The adoption of a 3D orientation is explained as well not by a metaphysical commitment but by practical considerations – it allows an easier integration of content from other ontologies.

In chapter 5 it is stated explicitly that "SUMO was constructed with reference to very pragmatic principles". This makes it "simpler to use than Cyc".

The SUMO and Semantic Web paper:

This paper mentions two "key advantages" of a standard upper ontology: First, such an ontology can serve as a design aid that alerts ontology designers to fundamental notions they should include in their ontologies. Second, the reuse of a standard upper ontology greatly facilitates ontology design in that the same fundamental notions do not have to be re-modelled with every new ontology. While these are "inherent advantages that the SUMO has as an upper-level ontology", the alignment of SUMO to WordNet[21] makes it particularly useful for natural language processing tasks.

In the remaining part of the paper two more advantages of SUMO are mentioned, namely the integration of domain ontologies (applications using different domain ontologies can interact at a more general semantic level if the domain ontologies are consistent with the same upper-level ontology) and the improvement of searching (by marking up documents with the concepts of an extensive common ontology).

The progress report:

In the introduction the broad purpose of the SUO effort (and hence also SUMO's) are explained. It "is targeted for use in automated inference, semantic interoperability between heterogeneous information systems and natural language processing applications".

In contrast to the second SUO candidate, the IFF Foundation Ontology, SUMO "embodies the view that we can construct a single, consistent, upper-level ontology that will be adequate for real applications".

The main part of the paper summarises the discussions from the SUO mailing list. Here we can see that the development of SUMO was guided by pragmatic decisions. We read that choices are made by evaluating the "cost and benefit tradeoff" of the options. Certain semiotics concepts were added to SUMO because of their relevance to reasoning in intellectual property issues and media. The notion of organisation was modeled to fit the needs voiced by various participants on the mailing list. SUMO also caters for agent communication by providing concepts for types of speech acts.

In the conclusion the benefit of SUMO to natural language processing after its mapping to WordNet is mentioned again. In addition, the possibility of navigating SUMO using natural language terms is an improvement of its usability.

The motivations for SUMO can be summed up as follows: The design of SUMO was

21 WordNet is a large lexical database of English. See http://wordnet.princeton.edu/

guided much more by pragmatic han by theoretic considerations. SUMO was designed in order to be used for engineering a wide range of "real" applications. Among these applications are specifically named the areas of web searching, automated inference, semantic interoperability between information systems, ecommerce, automated language understanding, and natural language processing (NLP); NLP being the most prominent field.

### 3.3 Motivations for DOLCE

Sweetening ontologies with DOLCE:

According to Gangemi et al. the main purpose of foundational ontologies is to allow for cooperation between artificial agents or between artificial and human agents by negotiating meaning. As one of the ontologies within the WonderWeb Library of Foundational Ontologies, DOLCE is intended to be used by "Semantic Web applications". The authors of the paper believe that DOLCE's clear cognitive bias "is very important for the Semantic Web (especially if we consider its intrinsically social nature [...])". Nevertheless, it is not intended as a candidate for a universal standard ontology but rather "to act as a starting point for comparing and elucidating the relationships with other future modules of the library".

DOLCE has a clear theoretical starting point. It models the world according to human language and cognition.

DOLCE's ontological choices, as for example the distinction between endurants and perdurants, are motivated by their previously selected approach.

Deliverable D18:

After a general introduction to foundational ontologies the reader is informed about the WonderWeb Foundational Ontologies Library (WFOL). The WFOL is presented as an important resource for Semantic Web applications. Masolo et al. "believe that the most important challenge for the Semantic Web is [...] the careful isolation of the fundamental ontological options and their formal relationships" (p. 3). The main goals of the WFOL are given as follows (p. 3-4):

- serve as a starting point for building new ontologies
- serve as a reference point for comparisons among different ontological

approaches

- serve as a common framework for analyzing, harmonizing and integrating existing ontologies and metadata standards.

DOLCE was built after "a core set of key ontological assumptions, focusing on the needs of other projects we were involved in," (p. 5) had been described.
Further statements about DOLCE are congruent with what is said by Gangemi et al. (2002).

Summing up the motivations for DOLCE, it is easy to say that DOLCE was intended to be a starting point for comparing different ontological choices and design options. Therefore it was modelled very carefully according to a certain world-view previously agreed upon. Only in its second stage, DOLCE was designed for the actual use in the Semantic Web. (The "needs of other projects" (Masolo et al. 2003, p. 5) that played a role in the construction of DOLCE are unfortunately not explained in more detail.)

### 3.4 Comparison

It is possible to make out three motivational differences between SUMO and DOLCE.
First, there are the developers' presuppositions about what kind of foundational ontologies are needed. The SUMO developers adopt the stance that there should be a single comprehensive standard ontology, while the DOLCE developers hold that there should be a library of different modules representing possible ontological choices in a well documented way.
Second, SUMO and DOLCE differ slightly in the use they are intended to be put to. Both are intended to be used in the Semantic Web, but while a lot of application areas are named for SUMO and the overall engineering context is stressed, is DOLCE being located in the context of a "social" Semantic Web.
Third, the approaches taken when developing SUMO and DOLCE differ considerably. The development of SUMO is strongly characterized by pragmatic choices. On the other hand the development of DOLCE is guided by thorough theoretic choices.

## 4. What are SUMO and DOLCE used for?

After the more theoretic presentation of SUMO and DOLCE in chapter 2, I analyse what SUMO and DOLCE have actually been used for by investigating their reception in the scientific community on the basis of citations.

### 4.1 Methodology

There are different ways to approach the question of who uses SUMO and DOLCE. The DOLCE website, for example, offers a list of institutions that "are using or have expressed interest in the DOLCE ontology" (LOA [n.d.]). The problem is that this list lists in most cases only an institution and a person. There is no additional information available on how the institution has been involved with DOLCE. Using this list as a starting point would require a lot of extra work. Without, one can not predict how complete and how up-to-date the list is.

Another way to find out more about the users and uses of SUMO and DOLCE is to look at the citations of the most relevant SUMO and DOLCE papers. While this approach would be more objective than a manually assembled list of users, it is on the other hand prone to exclude a certain group of users, namely those that used SUMO or DOLCE but did not write about it. This concerns most likely industrial and commercial projects.

Despite the drawbacks that each of the two approaches has, I decided to use the second approach. Analysing the citations promised to yield a relatively broad overview of the uses of SUMO and DOLCE while consuming less time resources. Furthermore it will allow for an objective comparison between SUMO and DOLCE.

Having decided on the approach, the next step was to determine the sources. The SUMO developers ask that "people working with SUMO cite [their] primary paper in any publications" (Pease 2009, August 11). So "Towards a standard upper ontology" by Niles & Pease (2001) lends itself to the purpose. Candidates for the DOLCE paper are the WonderWeb deliverables D15 and D18, and "Sweetening ontologies with DOLCE" by Gangemi et al. (2002), because they introduce the DOLCE ontology. The most citations by far[22] can be found for "Sweetening ontologies with DOLCE", so this paper will be used as the source for DOLCE.

22 As of October 12th, 2009, Google Scholar finds 409 citations for the paper by Gangemi et al., 138 for D18 and 6 for D15. Citesserx and ISI Web of Knowledge do not know the deliverables at all, but have 67 and 29 citations, respectively, for Gangemi et al.

Google Scholar documents 773 citations[23] for the primary SUMO paper. Citeseerx offers 120, and the ISI Web of Knowledge database does not know the paper.
I am going to use the citations provided by Google Scholar, because these include the Citeseerx citations.
Due to time constraints I have to limit the number of citations that I will analyse to 100.
How will I analyse the citations? Because it would take an inappropriate amount of time to read through all the articles citing the source papers, I need to define a criterion to decide beforehand which articles are relevant. Often an article only mentions SUMO or DOLCE as an example for a foundational ontology but does not discuss it further than that. To predetermine those publications that deal with SUMO or DOLCE in more detail, I consider only those that mention SUMO or DOLCE either in their titles or their abstracts. Although I might miss some publications that are about the usage of SUMO or DOLCE with this approach, I am confident that I will be able to find those that focus on either one of the ontologies.
For an initial analysis of the data thus retrieved I classify the articles broadly with respect to three categories: 1) field of origin (either academic or commercial context), 2) subject area, and 3) the way in which SUMO or DOLCE have been used.
Following the analysis I will sum up and compare the findings.

### 4.2 Analysis

Among the first 100 citations for the SUMO and for the DOLCE paper, I found twelve papers for SUMO and 14 papers for DOLCE.[24] After eliminating the two double entries among the DOLCE papers, I have to consider twelve citations for each of the source papers.
The results of the classification of the papers are shown in Table 3 and Table 4.

---

23 As of October 12$^{th}$, 2009. http://scholar.google.com/scholar?cites=12868820910366548094&hl=en
24 A list of these papers and their abstracts can be found in the appendix.

| | Document | Origin | Subject Area | Use of SUMO |
|---|---|---|---|---|
| 1 | Niles2003 | industry | information science | enhancement of SUMO |
| 2 | Stuckenschmidt 2004 | academia | information science | experimenting with SUMO |
| 3 | Atserias2004 | both | linguistics / information science | source of formal semantics |
| 4 | Ahrens2003 | academia | linguistics | source of formal semantics |
| 5 | Soldatova2006 | academia | topic oriented (theory of science) | foundation for domain ont. |
| 6 | Niles2004 | industry | information science | foundation for domain ont. |
| 7 | Farrar2002 | academia | linguistics | foundation for domain ont. |
| 8 | Black2006 | both | linguistics | source of formal semantics |
| 9 | Scheffczyk2006 | both | NLP | source of formal semantics |
| 10 | Lewis2001 | academia | linguistics | foundation for domain ont. |
| 11 | Mascardi2007 | academia | software agents | analysis of SUMO |
| 12 | Pease2004 | both | NLP | source of formal semantics |

*Table 3: The papers citing SUMO*

| | Document | Origin | Subject Area | Use of DOLCE |
|---|---|---|---|---|
| 1 | Gangemi2003a | academia | Semantic Web | foundation for domain ont. |
| 2 | Gangemi2003b | academia | linguistics / information science | source of formal semantics |
| 3 | Mika2004 | academia | Semantic Web / web services | foundation for domain ont. |
| 4 | Gangemi2003c | academia | topic oriented (law) | foundation for domain ont. |
| 5 | Sagri2003 | academia | topic oriented (law) | foundation for a content descr. model |
| 6 | Frank2006 | academia | philosophy of mind / information science | ontological study |
| 7 | Bateman2004 | academia | NLP / topic oriented (robotics, spatial ontology) | foundation for domain ont. |
| 8 | Guizzardi2004 | academia | topic oriented (economics) | foundation for found. ont. |
| 9 | Probst2006 | academia | topic oriented (geospatial information) | source of formal semantics / foundation for "conc. model" |
| 10 | Guizzardi2005 | academia | information science | foundation for found. ont. |
| 11 | Probst2007 | academia | topic oriented (geospatial information) | foundation for core ont. |
| 12 | Petridis2004 | academia | topic oriented (multimedia content) | foundation for domain ont. |

*Table 4: The papers citing DOLCE*

Classification after *origin* takes place according to the authors' affiliation as provided in the papers. All papers citing DOLCE are of academic origin. Also for SUMO, papers of academic origin make up the biggest part (one half). The other half are either papers of mixed (1/3) or of purely industrial origin (1/6). The fact that SUMO is cited by authors with an industrial background is certainly due to its own authors' industrial affiliation, and the more so as Pease and Niles, who are the authors of the primary source for SUMO, are the (co-)authors of many papers citing SUMO.

Classification after *subject area* was done according to the information from the abstracts. Assigning each paper to one subject area involved some difficulties, as many papers do not explicitly name their subject area, or are concerned with both information science and an individual subject area. The latter were classified according to their individual subject areas because I think it interesting to see in which disciplines SUMO and DOLCE are used. For the former I determine a subject area from the terms in the abstracts and, if available, the rest of the texts.

The finding with regard to subject areas is that while SUMO is mostly used or discussed in linguistics (5 papers from the field of linguistics, or 7 if including NLP[25]) and information science (4 papers), DOLCE is used or discussed in connection with a diverse range of subject areas: information science (3 papers), Semantic Web (2), law (2), robotics (1), geospatial information (2), economics (1), multimedia content (1).

Classification after how the authors of the citing papers had made *use* of SUMO or DOLCE is done on the basis of the papers' abstracts as well.

The majority of papers make use of the one or the other foundational ontology as either a foundation for a new ontology or as a resource of formally defined world knowledge. These two uses are not very different, as founding a domain ontology on a foundational ontology basically means that you employ its formally defined semantics. I keep the distinction even so, because it allows us to note that not every use of SUMO's or DOLCE's formal semantics is aimed at constructing an ontology. For SUMO we count 4 papers that use SUMO as foundation for a domain ontology, and 5 papers that exploit its formal semantics in another way. For DOLCE there are 10 papers using it as a foundation for a new domain, core or even foundational ontology[26]. 2 papers are using DOLCE's formal

25 Natural Language Processing (NLP) is concerned with how computers can process the language spoken or written by humans. It is a field of linguistics, but closely connected to computer science, too.

26 The uses "foundation for conceptual model" and "foundation for a content description model" are counted under this heading, too, because I figured that a conceptual model and a content descrition

semantics differently (one of them also counted in the former category).
Three papers about SUMO are citing it for fundamentally different purposes: One of them is using it to test their newly developed method, one is analysing several foundational ontologies, among them SUMO, and one is about the further development of SUMO itself. The only non-standard use of DOLCE is especially remarkable because that paper cites DOLCE for its theoretical background on which the paper's author bases further studies.

### 4.3 Findings and limitations

Although my study of citations is by no means representative, it yields some interesting preliminary results. We can make out a tendency in the uses of both SUMO and DOLCE. While SUMO is mostly used as a source of formal semantics in linguistics and NLP and only secondly as a foundation for domain ontologies, DOLCE serves as a foundation for a diverse range of ontologies in different subject areas. Considering the motivations for SUMO and DOLCE (chapter 3), the results for SUMO match SUMO's intended purpose. From its motivations we also expect SUMO to be often used in a commercial context, but the citation analysis does not allow us to draw conclusions in that context. (Commercial applications are usually not described in published papers.) As DOLCE was primarily intended to serve as a reference point with regard to ontological design options, it comes as a surprise that most publications in the citation study actually use DOLCE as the foundation for their own ontologies. It also comes as a surprise to me that the Descriptive Ontology for Linguistic and Cognitive Engineering is not used within linguistics and NLP more often. This field seems to be dominated by SUMO, and the question "why" remains open.
There is another flaw with choosing the citations from Google Scholar that should be mentioned. Google Scholar ranks its results mainly after the number of times they have been cited. This means that recent publications – that have not collected many citations yet – will never appear among the first results. This might explain why there are no 2008 or 2009 publications among my sample.

model would be very similar to domain ontologies. This is debatable, of course.

## 5. Conclusion

In this paper I compared SUMO and DOLCE.

DOLCE is a foundational ontology that is modelled according to human language and cognition. Therefore it is descriptive and multiplicative, and adopts possibilism as well as a 3D view. SUMO is a foundational ontology developed with an engineering oriented context in mind. It adopts a 3D view of the world, and tends towards descriptivism, multiplicativism and actualism; the latter three choices however being rather unclear. (compare chapter 2).

SUMO was developed as a single comprehensive standard ontology, to be used in a wide range of applications and the Semantic Web. Its development is strongly characterised by pragmatic choices. DOLCE on the other hand was intended to serve as a reference module in a library of different foundational ontologies. Its development was guided by thorough theoretic choices, and its possible application in the Semantic Web of minor importance. (compare chapter 3).

Both SUMO and DOLCE are put to use by several authors. Of their application in academic context it seems that while SUMO is mostly used as a source of formal semantics in linguistics and NLP, DOLCE serves as a foundation for a diverse range of ontologies in different subject areas. (compare chapter 4).

My research question "What use do foundational ontologies have?" has been answered from different perspectives. The definitions in the introduction communicated that ontologies are a means of organising information with enhanced abilities. They do not only help people to gain a common understanding of terms but are also machine-understandable. Foundational ontologies define very basic concepts and provide the grounds for the development of more specialised domain ontologies. The examples of SUMO and DOLCE have shown that there are different approaches to and different purposes behind developing a foundational ontology, and that, despite their unlike level of theoretic grounding, they both are used in several applications.

Looking into the future of library and information science, I assume that ontologies will grow in importance, at least to the same degree as the Semantic Web advances. This paper has provided a first insight into foundational ontologies. They are very complex structures. Before we can apply them in LIS we need to acquire a deeper understanding of them; which, in turn, requires background knowledge above all in logics, but also in computer science and philosophy. Beyond that we should look at successful applications in order to determine which foundational ontology is best used for what. Studies that directly compare the use of different ontologies or the choice of different basic ontological modelling approaches would be valuable.

## Bibliography


Angele, J., & Lausen, G. (2004). Ontologies in F-logic. In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 29-50). Berlin, Heidelberg: Springer.

Antoniou, G., & Harmelen, F. V. (2004). Web Ontology Language: OWL. In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 67-92). Berlin, Heidelberg: Springer.

Baader, F., Horrocks, I., & Sattler, U. (2004). Description Logics. In *Handbook on Ontologies* (pp. 3-28). Berlin, Heidelberg: Springer.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5). Retrieved from http://www.scientificamerican.com/article.cfm?id=the-semantic-web

Borgo, S., Gangemi, A., Guarino, N., Masolo, C., & Oltramari, A. (2002). *WonderWeb Deliverable D15 : Ontology RoadMap*. Padova, Italy: ISTC-CNR. Retrieved November 2, 2009, from http://wonderweb.man.ac.uk/deliverables/documents/D15.pdf

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies: where is their meeting point? *Data & Knowledge Engineering*, *46*(1), 41-64.

Euzenat, J., & Shvaiko, P. (2007). *Ontology Matching*. Berlin, Heidelberg: Springer.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. (2002). Sweetening Ontologies with DOLCE. In Asunción Gómez-Pérez & V. Richard Benjamins (Eds.), *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, Lecture Notes in Computer Science (pp. 223-233). Berlin, Heidelberg: Springer. Retrieved July 5, 2009, from http://dx.doi.org/10.1007/3-540-45810-7_18

Genesereth, M. R., & Richard E. Fikes. (1992). *Knowledge Interchange Format : Version 3.0 Reference Manual*. Stanford, California: Computer Science Department, Stanford University. Retrieved November 2, 2009, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.47.1022&rep=rep1&type=pdf

Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, *5*(2), 199-220.

Gruber, T. (2007). Ontology (Computer Science) - definition in Encyclopedia of Database Systems. Retrieved November 2, 2009, from http://tomgruber.org/writing/ontology-definition-2007.htm

Guarino, N. (1998). Formal Ontology and Information Systems. In *Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998* (pp. 3-15). Presented at the Formal Ontology in Information Systems, Amsterdam: IOS Press. Retrieved November 2, 2009, from http://www.loa-cnr.it/Papers/FOIS98.pdf

Guarino, N., & Giaretta, P. (1995). Ontologies and knowledge bases: Towards a terminological clarification. In N. Mars (Ed.), *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing* (pp. 25-32). Amsterdam: IOS Press. Retrieved November 2, 2009, from www.loa-cnr.it/Papers/KBKS95.pdf

Guarino, N., & Welty, C. (2002). Evaluating ontological decisions with OntoClean. *Communications of the ACM*, *45*(2), 61-65.

Hügli, A., & Lübcke, P. (Eds.). (2001). *Philosophielexikon : Personen und Begriffe der abendländischen Philosophie von der Antike bis zur Gegenwart*. rowohlts enzyklopädie (4th ed.). Reinbek bei Hamburg: Rowohlt.

IEEE. (2003, December 28). Standard Upper Ontology Working Group (SUO WG) - Home Page. *Standard Upper Ontology Working Group (SUO WG)*. Retrieved November 2, 2009, from http://suo.ieee.org/

LOA. (n.d.). Laboratory for Applied Ontology - DOLCE. *ISTC - CNR, Laboratory for Applied Ontology (LOA)*. Retrieved November 2, 2009, from http://www.loa-cnr.it/dolcus.html

Mascardi, V., Cordì, V., & Rosso, P. (2007). A comparison of upper ontologies. In *Proc. Conf. on Agenti e industria: Applicazioni tecnologiche degli agenti software, WOA07, Genova, Italy*. Retrieved November 2, 2009, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.144.1792&rep=rep1&type=pdf

McBride, B. (2004). The Resource Description Framework (RDF) and its Vocabulary Description Language RDFS. In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 51-66). Berlin, Heidelberg: Springer.

Niles, I., & Pease, A. (2001). Towards a Standard Upper Ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems - Volume 2001* (pp. 2-9). Ogunquit, Maine, USA: ACM. Retrieved November 2, 2009, from http://doi.acm.org/10.1145/505168.505170

Oberle, D. (2006). *Semantic Management of Middleware*. New York, NY, USA: Springer.

Oberle, D., Ankolekar, A., Hitzler, P., Cimiano, P., Sintek, M., Kiesel, M., et al. (2007). DOLCE ergo SUMO: On foundational and domain models in the SmartWeb Integrated Ontology (SWIntO). *Web Semantics: Science, Services and Agents on the World Wide Web*, *5*(3), 156-174.

Passin, T. B. (2004). *Explorer's Guide to the Semantic Web*. Greenwich, CT, USA: Manning.

Pease, A. (2008, August 22). Ontology Portal - Publications. *The Suggested Upper Merged Ontology (SUMO) - Ontology Portal*. Retrieved November 2, 2009, from http://www.ontologyportal.org/Pubs.html

Pease, A. (2008, November). *SUMO (Suggested Upper Merged Ontology)*. New York, NY, USA: Institute of Electrical and Electronics Engineers, Inc. Retrieved November 2, 2009, from http://sigmakee.cvs.sourceforge.net/*checkout*/sigmakee/KBs/Merge.kif

Pease, A. (2009, August 11). The Suggested Upper Merged Ontology (SUMO) - Ontology Portal. *The Suggested Upper Merged Ontology (SUMO) - Ontology Portal*. Retrieved November 2, 2009, from http://www.ontologyportal.org/

Pease, A., & Niles, I. (2002). IEEE standard upper ontology: a progress report. *The Knowledge Engineering Review*, *17*(1), 65-70.

Pease, A., Niles, I., & Li, J. (2002). The Suggested Upper Merged Ontology: a large ontology for the semantic web and its applications. In *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*. Edmonton, Canada. Retrieved November 2, 2009, from http://home.earthlink.net/~adampease/professional/Pease.ps

Schreiber, G. (2008). Knowledge Engineering. In F. V. Harmelen, V. Lifschitz, & B. Porter (Eds.), *Handbook of Knowledge Representation*, Foundations of artificial intelligence (Vol. 1, pp. 929-946). Amsterdam: Elsevier.

Semy, S. K., Pulvermacher, M. K., & Obrst, L. J. (2004). *Toward the Use of an Upper Ontology for U.S. Government and U.S. Military Domains: An Evaluation* (Technical report No. MTR 04B0000063). Bedford, Massachusetts: MITRE Corporation. Retrieved November 2, 2009, from http://handle.dtic.mil/100.2/ADA459575

Staab, S., & Studer, R. (Eds.). (2004). *Handbook on Ontologies*. International Handbooks on Information Systems. Berlin, Heidelberg: Springer.

Stock, W. G., & Stock, M. (2008). *Wissensrepräsentation : Informationen auswerten und bereitstellen*. München: Oldenbourg.

Stuckenschmidt, H. (2009). *Ontologien : Konzepte, Technologien und Anwendungen*. Informatik im Fokus. Berlin, Heidelberg: Springer.

## Appendix: Citations for SUMO and DOLCE

### Citations for DOLCE (Gangemi et al. 2002), retrieved by Google Scholar on October 12, 2009. Abstracts added.

Bateman2004:

Bateman, J., & Farrar, S. (2004). Towards a generic foundation for spatial ontology. In A. C. Varzi & L. Vieu (Eds.), *Formal Ontology in Information Systems: Proceedings of the Third International Conference (FOIS 2004)*. Amsterdam: IOS Press.

Abstract: In this paper we describe the results of our current efforts to establish a foundation for combining spatial ontologies. We draw on the state of the art in computational ontologies engineering, ontologies for geographic information systems and qualitative reasoning in order to build a framework within which these, and other, approaches may be placed. The skeleton of the framework is provided by the DOLCE ontology augmented by several components of Smith's BFO. We point out the particular importance of including notions of variable granularity, layers and sites in order to do justice to the requirements of practical ontological applications. The target applications for our own work are spatial reasoning for robotic systems and natural language interaction.

Frank2006:

Frank, A. U. (2006). Distinctions Produce a Taxonomic Lattice: Are These the Units of Mentalese? In B. Bennett & C. Fellbaum (Eds.), *Formal Ontology in Information Systems: Proceedings of the Fourth International Conference (FOIS 2006)* (pp. 27-38). Amsterdam: IOS Press.

Abstract: Ontologies describe a conceptualization of a part of the world relevant to some application. What are the units of conceptualizations? Current ontologies often equate concepts with words from natural languages. Words are certainly not the smallest units of conceptualization, neither are the sets of synonyms of WordNet or other linguistically justified units. I suggest to take distinctions as basic units and to construct concepts from them whereas other approaches start with concepts and discover properties that distinguish them. Distinctions separate concepts and produce a taxonomic lattice, which contains the named concepts together with other potential conceptual units. The taxa are organized in a superclass/subclass (better supertaxa/subtaxa) relation and for any two taxa there is always a single least common supertaxon. Algorithms to maintain such a taxonomic structure and methods to combine different taxonomies are shown, using a four valued (relevance) logic as introduced by Belnap [1]. The novel aspect of the method is that distinctions that are only meaningful in the context of other distinctions restrict the lattice of concepts to the meaningful subset.
The approach is restricted to the is_a relation between classes; it relates to Formal Concept Analysis, but replaces the "formal attributes" with (necessary) distinctions and uses a four-valued logic. It stresses the focus of recent ontological studies like DOLCE or WonderWeb on qualities; it is expected that distinctions as introduced here for the is_a hierarchy influence the mereological aspects of an ontology (i.e., the part_of relation) and connect to Gibson's affordances [2] and contribute to the classification of operations.

[Gangemi2003d:

Gangemi, A., Navigli, R., & Velardi, P. (2003). Axiomatizing WordNet glosses in the OntoWordNet project. In *Workshop on Human Language Technology for the Semantic Web and Web Services* (pp. 51-76).

Abstract: In this paper we present a progress report of the OntoWordNet project, a research program aimed at achieving a formal specification of WordNet. Within this program, we developed a hybrid bottom-up top-down methodology to automatically extract association relations from WordNet, and to interpret those associations in terms of a set of conceptual relations, formally defined in the DOLCE foundational ontology. Preliminary results provide us with the conviction that a research program aiming to obtain a consistent, modularized, and axiomatized ontology from WordNet can be completed in acceptable time with the support of semi-automatic techniques.]

Gangemi2003a:
Gangemi, A., & Mika, P. (2003). Understanding the Semantic Web through Descriptions and Situations. In R. Meersman, Z. Tari, & D. C. Schmidt (Eds.), *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, Lecture Notes in Computer Science (pp. 689-706). Berlin, Heidelberg: Springer.

Abstract: The Semantic Web is a powerful vision that is getting to grips with the challenge of providing more human-oriented web services. Hence, reasoning with and across distributed, partially implicit assumptions (contextual knowledge), is a milestone. Ontologies are a primary means to deploy the Semantic Web vision, but few work has been done on them to manage the context-dependency of Web knowledge. In this paper we introduce an ontology for representing a variety of reified contexts and states of affairs, called D&S, currently implemented as a plug-in to the DOLCE foundational ontology, and its application to two cases: an ontology for communication situations and roles, and an ontology for peer-to-peer communication. The reified contexts represented in D&S have a rich structure, and are a middleware between full-fledged formal contexts and theories, and the often poor vocabularies implemented in Web ontologies...

Gangemi2003b:
Gangemi, A., Navigli, R., & Velardi, P. (2003). The OntoWordNet Project: Extension and Axiomatization of Conceptual Relations in WordNet. In R. Meersman, Z. Tari, & D. C. Schmidt (Eds.), *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, Lecture Notes in Computer Science (pp. 820-838). Berlin, Heidelberg: Springer.

Abstract: In this paper we present a progress report of the OntoWordNet project, a research program aimed at achieving a formal specification of WordNet. Within this program, we developed a hybrid bottom-up top-down methodology to automatically extract association relations from WordNet, and to interpret those associations in terms of a set of conceptual relations, formally defined in the DOLCE foundational ontology. Preliminary results provide us with the conviction that a research program aiming to obtain a consistent, modularized, and axiomatized ontology from WordNet can be completed in acceptable time with the support of semi-automatic techniques.

Gangemi2003c:
Gangemi, A., Prisco, A., Sagri, M., Steve, G., & Tiscornia, D. (2003). Some Ontological Tools to Support Legal Regulatory Compliance, with a Case Study. In *On The Move*

*to Meaningful Internet Systems 2003: OTM 2003 Workshops* (pp. 607-620).

Abstract: The increasing development of legal ontologies seems to offer satisfactory solutions to legal knowledge formalization, which in past experiences lead to a limited exploitation of legal expert systems for practical and commercial use. The paper describes some ontology-based tools that enable legal knowledge formalization. Jurwordnet is an extension to the legal domain of the Italian version of EuroWordNet. It is a content description model for legal information and a lexical resource for accessing multilingual and heterogeneous information sources. Its concepts are organised according to a Core Legal Ontology (CLO), based on DOLCE+, an extension of the DOLCE foundational ontology. Jurwordnet and CLO are also used to represent the assessment of legal regulatory compliance across different legal systems or between norms and cases. An example is discussed concerning compliance between EC directives and national legislations.

Guizzardi2004:
Guizzardi, G., & Wagner, G. (2004). A unified foundational ontology and some applications of it in business modeling. In *Workshop on Enterprise Modeling and Ontologies for Interoperability, 16th International Conference on Advanced Information Systems Engineering (CAISE), Riga*.

Abstract: Foundational ontologies provide the basic concepts upon which any domain-specific ontology is built. This paper presents a new foundational ontology, UFO, and shows how it can be used as a guideline in business modeling and for evaluating business modeling methods. UFO is derived from a synthesis of two other foundational ontologies, GFO/GOL and OntoClean/DOLCE. While their main areas of application are natural sciences and linguistics/cognitive engineering, respectively, the main purpose of UFO is to provide a foundation for conceptual modeling, including business modeling.

[Guizzardi2005a:
Guizzardi, G., & Wagner, G. (2005). Some Applications of a Unified Foundational Ontology in Business Modeling. *Business Systems Analysis with Ontologies*, 345–367.

Abstract: Foundational ontologies provide the basic concepts upon which any domain-specific ontology is built. This paper presents a new foundational ontology, UFO, and shows how it can be used as a guideline in business modeling and for evaluating business modeling methods. UFO is derived from a synthesis of two other foundational ontologies, GFO/GOL and OntoClean/DOLCE. While their main areas of application are natural sciences and linguistics/cognitive engineering, respectively, the main purpose of UFO is to provide a foundation for conceptual modeling, including business modeling.]

Guizzardi2005:
Guizzardi, G., & Wagner, G. (2005). Towards Ontological Foundations for Agent Modelling Concepts Using the Unified Fundational Ontology (UFO). In P. Bresciani, P. Giorgini, B. Henderson-Sellers, G. Low, & M. Winikoff (Eds.), *Agent-Oriented Information Systems II*, Lecture Notes in Computer Science (pp. 110-124). Berlin, Heidelberg: Springer.

Abstract: Foundational ontologies provide the basic concepts upon which any domain-specific ontology is built. This paper presents a new foundational ontology, UFO, and shows how it can be used as a foundation of agent concepts and for evaluating agent-oriented modelling methods. UFO is derived from a synthesis of two other foundational ontologies, GFO/GOL and OntoClean/DOLCE. While their main areas of application are the natural sciences and linguistics/cognitive engineering, respectively, the main purpose of UFO is to provide a foundation for conceptual modelling, including agent-oriented modelling.

Mika2004:
Mika, P., Oberle, D., Gangemi, A., & Sabou, M. (2004). Foundations for service ontologies: aligning OWL-S to dolce. In *Proceedings of the 13th international conference on World Wide Web* (pp. 563-572). New York, NY, USA: ACM.

Abstract: Clarity in semantics and a rich formalization of this semantics are important requirements for ontologies designed to be deployed in large-scale, open, distributed systems such as the envisioned Semantic Web. This is especially important for the description of Web Services, which should enable complex tasks involving multiple agents. As one of the first initiatives of the Semantic Webcommunity for describing Web Services, OWL-S attracts a lot of interest even though it is still under development. We identify problematic aspects of OWL-S and suggest enhancements through alignment to a foundational ontology. Another contribution of our work is the Core Ontology of Services that tries to fill the epistemological gap between the foundational ontology and OWL-S. It can be reused to align other Web Service description languages as well. Finally, we demonstrate the applicability of our work by aligning OWL-S' standard example called CongoBuy.

Petridis2004:
Petridis, K., Kompatsiaris, I., Strintzis, M. G., Bloehdorn, S., Handschuh, S., Staab, S., et al. (2004). Knowledge representation for semantic multimedia content analysis and reasoning. In *Proceedings of the European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT)*. London.

Abstract: In this paper, a knowledge representation infrastructure for semantic multimedia content analysis and reasoning is presented. This is one of the major objectives of the aceMedia Integrated Project where ontologies are being extended and enriched to include low-level audiovisual features, descriptors and behavioural models in order to support automatic content annotation. More specifically, the developed infrastructure consists of the core ontology based on extensions of the DOLCE core ontology and the multimedia-specific infrastructure components. These are, the Visual Descriptors Ontology, which is based on an RDFS representation of the MPEG-7 Visual Descriptors and the Multimedia Structure Ontology, based on the MPEG-7 MDS. Furthermore, the developed Visual Descriptor Extraction tool is presented, which will support the initialization of domain ontologies with multimedia features.

Probst2007:
Probst, F. (2007). *Semantic reference systems for observations and measurements*. (Doctoral thesis). Westfälische Wilhelms - Universität Münster.

Abstract: What is needed to enable communication about observation and measurement results in information systems? Information system ontologies make a certain

conceptualization explicit and partially account for the meanings of symbols associated with this conceptualization. Yet, the meaning of symbols denoting observation results such as /100m/, /328ft/ or /deep/ cannot be specified with currently available foundational ontologies. They fail to separate the ontological nature of an observable quality from the specification of how to observe and name the observation result. In this thesis, we employ and extend the foundational ontology DOLCE for making the meaning of symbols explicit that denote observation and measurement results. This involves dealing with ontological questions like “What kinds of observable qualities exist, in which entity does the observed quality inhere and how are the magnitudes of the observed quality structured?”
Then, in order to capture the semantic aspects of an observation result, we introduce semantic
reference spaces, dealing with semantic questions like “Do the symbols /100m/, /328 feet/ or
/deep/ have the same meaning? Do these symbols refer to the same entity, e.g. the depth magnitude of a lake? How to establish a unit of measure?” We posit that only if agreement is reached on the ontological questions, the semantic questions can be approached effectively. Extending the foundational ontology DOLCE with an ontology for spatial qualities, an ontology for quality spaces, and an ontology for semantic reference spaces leads to a theory of semantic reference systems that allows for specifying and comparing the meaning of symbols denoting observation results.

Probst2006:

Probst, F. (2006). Ontological Analysis of Observations and Measurements. In M. Raubal (Ed.), *Geographic, Information Science* (pp. 304-320). Berlin, Heidelberg: Springer.

Abstract: Geographic information is based on observations or measurements. The Open Geospatial Consortium (OGC) has developed an implementation specification for observations and measurements (O&M). It specifies precisely how to encode information. Yet, the O&M conceptual model does not specify precisely which real-world entities are denoted by the specified information objects. We provide formal semantics for the central O&M terms by aligning them to the foundational ontology DOLCE. The alignment to a foundational ontology
restricts the possible interpretations of the central elements in the O&M model and establishes explicit relations between categories of real world entities and classes of information objects. These relations are essential for assessing semantic interoperability between geospatial information sources.

Sagri2003:

Sagri, M., & Tiscornia, D. (2003). Metadata for content description in legal information. In *Database and Expert Systems Applications, 2003. Proceedings. 14th International Workshop on* (pp. 745-749).

Abstract: The subject of this paper is a description of the Jur-Wordnet (Jur-IWN) project, an extension for legal domain of the Italian version of EuroWordNet database, linked to the Interlingual Index (ILI) records, The data base aims at providing both a content description model for legal information and a tool for accessing multilingual and heterogeneous information sources. The double perspective of application of the database involves a double interpretation of the model, and of its features. Conceived as a lexical source, terms are linked to each other throughout semantic relation, while, as a content description model, the concepts are organised according to stronger assumptions about the ontological nature of entities that populate the legal domain and about their relationship. The crucial problem is to save the global consistency of the model, in the view of the interchange between their components the lexical view need a clear definition of the boundaries between common language and technical legal terminology, while the conceptualisation of the core entities need to be linked in coherent way with upper level categories and the lexical 'leafs'. [...]

**Citations for SUMO (Niles & Pease 2001),**
**retrieved by Google Scholar on October 12, 2009. Abstracts added.**

Ahrens2003:

Ahrens, K., Chung, S. F., & Huang, C. R. (2003). Conceptual Metaphors: Ontology-based representation and corpora driven Mapping Principles. In *Proceedings of the ACL Workshop on the Lexicon and Figurative Language* (pp. 35–41). Sapporo, Japan: Association for Computational Linguistics.

Abstract: The goal of this paper is to integrate the Conceptual Mapping Model with an ontology-based knowledge representation (i.e. Suggested Upper Merged Ontology (SUMO)) in order to demonstrate thatconceptual metaphor analysis can be restricted and eventually, automated. In particular, we will propose a corporabased operational definition for Mapping Principles, which are explanations of why a conventional conceptual metaphor has a particular source-target domain pairing. This paper will examine 2000 random examples of ‘economy’ (jingji) in Mandarin Chinese and postulate Mapping Principles based frequency and delimited with SUMO.

Atserias2004:

Atserias, J., Villarejo, L., Rigau, G., Agirre, E., Carroll, J., Magnini, B., et al. (2004). The meaning multilingual central repository. In *2nd International Global Wordnet Conference, January 20-23, 2004: proceedings, 23-30*. Brno, Czech Republic: Masaryk University.

Abstract: This paper describes the first version of the Multilingual Central Repository, a lexical knowledge base developed in the framework of the MEANING project. Currently the Mcr integrates into the EuroWordNet framework five local wordnets (including four versions of the English WordNet from Princeton), an upgraded version of the EuroWordNet Top Concept ontology, the MultiWordNet Domains, the Suggested Upper Merged Ontology (SUMO) and hundreds of thousand of new semantic relations and properties automatically acquired from corpora. We believe that the resulting Mcr will be the largest and richest Multilingual Lexical Knowledge Base in existence.

Black2006:

Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., et al. (2006). Introducing the Arabic WordNet Project. In P. Sojka, K. Choi, C. Fellbaum, & P. Vossen (Eds.), *Proceedings of the third International WordNet Conference (GWC-06)*. South Jeju Island, Korea: Masaryk University, Brno.

Abstract: Arabic is the official language of hundreds of millions of people in twenty Middle East and northern African countries, and is the religious language of all Muslims of various ethnicities around the world. Surprisingly little has been done in the field of computerised language and lexical resources. It is therefore motivating to develop an Arabic (WordNet) lexical resource that discovers the richness of Arabic as described in Elkateb (2005). This paper describes our approach towards building a lexical resource in Standard Arabic. Arabic WordNet (AWN) will be based on the design and contents of the universally accepted Princeton WordNet (PWN) and will be mappable straightforwardly onto PWN 2.0 and EuroWordNet (EWN), enabling translation on the lexical level to English and dozens of other languages. Several tools specific to this task will be developed. AWN will be a linguistic resource with a deep formal semantic foundation. Besides the standard wordnet representation of senses, word
meanings are defined with a machine understandable semantics in first order logic. The basis for this semantics is the Suggested Upper Merged Ontology (SUMO) and its associated domain ontologies. We will greatly extend the ontology and its set of mappings to provide formal terms and definitions equivalent to each synset.

Farrar2002:
Farrar, S., Lewis, W., & Langendoen, T. (2002). A common ontology for linguistic concepts. In *Proceedings of the Knowledge Technologies Conference* (pp. 10-13). Austin, TX, USA.

Abstract: As part of a project called Electronic Metastructure for Endangered Languages Data 1 (EMELD), we have developed an ontology of concepts that encompasses a wide range of linguistic phenomena. The idea was initially conceived to facilitate both the knowledge sharing of annotated linguistic data and the searching of disparate language corpora. Such an ontology, however, is needed outside of the EMELD project for enhancing performance of the Semantic Web, for developing expert systems capable of linguistic analysis, and for providing a theory-neutral backbone in the processing of scientific documents pertaining to the linguistics domain. With an eye toward acceptance by the knowledge engineering community in general, we built the linguistic ontology on top of the Standard Upper Merged Ontology (SUMO).

Lewis2001:
Lewis, W., Farrar, S., & Langendoen, D. T. (2001). Building a knowledge base of morphosyntactic terminology. In S. Bird, P. Buneman, & M. Liberman (Eds.), *Proceedings of the IRCS workshop on linguistic databases* (pp. 150–156). Pennsylvania, Philadelphia, USA: University of Pennsylvania.

Abstract: This paper describes the beginning of an effort within the Linguist List's Electronic Metastructure for Endangered Languages Data (E-MELD) project to develop markup recommendations for representing the morphosyntactic structures of the world's endangered languages. Rather than proposing specific markup recommendations as in the Text Encoding Initiative (TEI), we propose to construct an environment for comparing data sets using possibly different markup schemes. The central feature of our proposed environment is an ontology of morphosyntactic terms with multiple inheritance and a variety of relations holding among the terms. We are developing our ontology using the Protégé editor, and are extending an existing upper-level ontology known as SUMO.

Mascardi2007:
Mascardi, V., Cordì, V., & Rosso, P. (2007). A comparison of upper ontologies. In *Proc. Conf. on Agenti e industria: Applicazioni tecnologiche degli agenti software, WOA07, Genova, Italy*.

Abstract: Upper Ontologies are quickly becoming a key technology for integrating heterogeneous knowledge coming from different sources. In fact, they may be exploited as a “lingua franca” by intelligent software agents in all those scenarios where it is impossible (or there is no will) for an agent to disclose its own entire ontology to other agent, despite the need to communicate with it. This paper represents the very preliminary step towards the exploitation of Upper Ontologies as bridges for allowing intelligent software agents to align heterogeneous ontologies in an automatic way, where we analyse the most up-to-date state-of-the-art. In this paper we analyse 7 Upper Ontologies, namely BFO, Cyc, DOLCE, GFO, PROTON, Sowa’s ontology, and SUMO, according to a set of standard software engineering criteria, and we synthesise our analysis in form of a comparative table. A summary of some existing comparisons drawn among subsets of the 7 Upper Ontologies that we deal with in this document, is also provided.

Niles2003:
Niles, I., & Pease, A. (2003). Linking lexicons and ontologies: Mapping wordnet to the suggested upper merged ontology. In H. R. Arabnia (Ed.), *Proceedings of the International Conference on Information and Knowledge Engineering* (pp. 412–416). Las Vegas, Nevada, USA: CSREA Press.

Abstract: Ontologies are becoming extremely useful tools for sophisticated software engineering. Designing applications, databases, and knowledge bases with reference to a common ontology can mean shorter development cycles, easier and faster integration with other software and content, and a more scalable product. Although ontologies are a very promising solution to some of the most pressing problems that confront software engineering, they also raise some issues and difficulties of their own. Consider, for example, the questions below:
• How can a formal ontology be used effectively by those who lack extensive training in logic
and mathematics?
• How can an ontology be used automatically by applications (e.g. Information Retrieval and Natural Language Processing applications) that process free text?
• How can we know when an ontology is complete?
In this paper we will begin by describing the upper-level ontology SUMO (Suggested Upper Merged Ontology), which has been proposed as the initial version of an eventual Standard Upper Ontology (SUO). We will then describe the popular, free, and structured WordNet lexical database. After this preliminary discussion, we will describe the methodology that we are using to align WordNet with the SUMO. We close this paper by discussing how this alignment of WordNet with SUMO will provide answers to the questions posed above.

Niles2004:
Niles, I., & Terry, A. (2004). The MILO: A general-purpose, mid-level ontology. In H. R.

Arabnia (Ed.), *Proceedings of the International Conference on Information and Knowledge Engineering* (pp. 5-19). Las Vegas, Nevada, USA: CSREA Press.

Abstract: The MILO (MId-Level Ontology) is a general-purpose mid-level ontology that reuses concepts from the freely available SUMO upper-level ontology. The purpose of the MILO is to provide deeper conceptual support to the various domain ontologies that have been developed under the SUMO. This paper describes the content and structure of the MILO, the methodology used in constructing this ontology, and the kinds of applications that can benefit from this ontology. It will be seen that the MILO constitutes a very useful bridge between the abstract content of the SUMO and the application-specific content of the domain ontologies.

Pease2004:
Pease, A., & Fellbaum, C. (2004). Language to logic translation with phrasebank. In P. Sojka, K. Pala, P. Smrz, C. Fellbaum, & P. Vossen (Eds.), *Proceedings of the Second International WordNet Conference (GWC 2004)* (pp. 187-192). Brno, Czech Republic: Masaryk University Brno.

Abstract: We discuss a restricted natural language understanding system and a proposed extension to it, which is a corpus of phrases. The Controlled English to Logic Translation (CELT) system allows users to make statements in a domain-independent, restricted English grammar that have a clear formal semantics and that are amenable to machine processing. CELT needs a large amount of linguistic and semantic knowledge. It is currently coupled with the Suggested Upper Merged Ontology, which has been mapped by hand to WordNet 1.6. We propose work on a new corpus of phrases (called PhraseBank) to be added to WordNet and linked to SUMO, which will catalog common English phrase forms, and their deep meaning in terms of the formal ontology. This addition should significantly expand the coverage and usefulness of CELT.

Scheffczyk2006:
Scheffczyk, J., Pease, A., & Ellsworth, M. (2006). Linking FrameNet to the suggested upper merged ontology. In B. Bennett & C. Fellbaum (Eds.), *Proceedings of Formal Ontology in Information Systems (FOIS-2006)* (Vol. 150, p. 289). Amsterdam: IOS Press.

Abstract: Deductive reasoning with natural language requires combining lexical resources with the world knowledge provided by ontologies. In this paper we describe the connection of FrameNet – a lexicon for English – to the Suggested Upper Merged Ontology (SUMO). We express general-domain links between FrameNet Semantic Types (ST) and SUMO classes in SUO-KIF, the language of SUMO. Based on these links, we have developed a semi-automatic, domain-specific approach for linking FrameNet Frame Elements (FE) to SUMO classes that is based on typical fillers of an FE in a particular domain. We thus provide restricted, ontology-based types on the fillers of FEs. Our work will enable several lines of experimentation for semantic parsing and ontology lexicalization.

Soldatova2006:
Soldatova, L. N., & King, R. D. (2006). An ontology of scientific experiments. *Journal of The Royal Society Interface, 3*(11), 795-803.

Abstract: The formal description of experiments for efficient analysis, annotation and sharing of results is a fundamental part of the practice of science. Ontologies are required to achieve this objective. A few subject-specific ontologies of experiments currently exist. However, despite the unity of scientific experimentation, no general ontology of experiments exists. We propose the ontology EXPO to meet this need. EXPO links the SUMO (the Suggested Upper Merged Ontology) with subject-specific ontologies of experiments by formalizing the generic concepts of experimental design, methodology and results representation. EXPO is expressed in the W3C standard ontology language OWL-DL. We demonstrate the utility of EXPO and its ability to describe different experimental domains, by applying it to two experiments: one in high-energy physics and the other in phylogenetics. The use of EXPO made the goals and structure of these experiments more explicit, revealed ambiguities, and highlighted an unexpected similarity. We conclude that, EXPO is of general value in describing experiments and a step towards the formalization of science.

Stuckenschmidt2004:

Stuckenschmidt, H., & Klein, M. (2004). Structure-Based Partitioning of Large Concept Hierarchies. In *The Semantic Web – ISWC 2004*, Lecture Notes in Computer Science (pp. 289-303). Berlin, Heidelberg: Springer.

Abstract: The increasing awareness of the benefits of ontologies for information processing has lead to the creation of a number of large ontologies about real-world domains. The size of these ontologies and their monolithic character cause serious problems in handling them. In other areas, e.g. software engineering, these problems are tackled by partitioning monolithic entities into sets of meaningful and mostly self-contained modules. In this paper, we suggest a similar approach for ontologies. We propose a method for automatically partitioning large ontologies into smaller modules based on the structure of the class hierarchy. We show that the structure-based method performs surprisingly well on real-world ontologies. We support this claim by experiments carried out on real-world ontologies including SUMO and the NCI cancer ontology. The results of these experiments are available online at http://swserver.cs.vu.nl/partitioning/.

Eigenständigkeitserklärung

Hiermit versichere ich, diese Arbeit eigenständig angefertigt und dabei keine anderen Quellen und Hilfsmittel als die angegebenen verwendet zu haben.